



De la collecte d'information à la restitution, comment utiliser Internet au cours de sa recherche ?

Bertrand Lefebvre, Sébastien Oliveau

► To cite this version:

Bertrand Lefebvre, Sébastien Oliveau. De la collecte d'information à la restitution, comment utiliser Internet au cours de sa recherche ?. Julie Beaujard. La restitution des données dans la recherche en sciences sociales : techniques et enjeux, pp.42-52, 2006, <<http://ajei.hypotheses.org/439>>. <hal-01140463>

HAL Id: hal-01140463

<https://hal.archives-ouvertes.fr/hal-01140463>

Submitted on 8 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ateliers Jeunes Chercheurs en Sciences Sociales de L'AJEI
8e édition, New Delhi, 28 février-5 mars 2005
La restitution des données dans la recherche en sciences sociales :
techniques et enjeux

in : Chapitre I / Techniques et supports de restitution des données

De la collecte d'information à la restitution, comment utiliser Internet au cours de sa recherche ?

Bertrand Lefebvre¹⁵

bertrand.lefebvre@csh-delhi.com

Sébastien Oliveau¹⁶

oliveau@parisgeo.cnrs.fr

L'atelier avait pour but de sensibiliser les jeunes chercheurs à l'utilisation efficace d'Internet pour sa recherche. On y a différencié la mise en valeur de ses travaux (abordé ici dans la seconde partie) de la recherche d'informations, que l'on envisagera tout de suite ici. Notre exposé est illustré de références à différents sites Web, que nous utilisons nous même dans notre activité de chercheurs indianistes et géographes.

A. La recherche d'information

Dans le cadre scientifique qui est le notre, il convient de distinguer deux sources d'informations disponibles sur Internet, les sources scientifiques (c'est-à-dire dont la qualité est garantie par un comité de lecture) des sources non-scientifiques (dont la qualité n'est donc pas contrôlée). On devra envisager aussi un certain nombre de sites qui, sans être garantis par un comité de lecture, émanent d'instances scientifiques, et que l'on nommera parascientifiques.

1. L'information non scientifique

L'information non scientifique est majoritaire sur Internet. On ne peut pas ici présenter toutes les sources disponibles sur Internet, on rappellera donc les plus importantes.

a) Les moteurs de recherche

Les moteurs de recherche (www.google.fr, www.yahoo.fr, etc.) constituent l'entrée la plus évidente... et la plus dangereuse dans l'Internet. Le principe en est simple, on tape des mots-clefs, et le moteur renvoie vers des sites qu'il juge pertinents en rapport avec ces mots-clefs. C'est d'ailleurs là que réside le principal problème : les critères de sélection des sites sont complètement opaques. La « pertinence » des résultats ne correspond pas à la pertinence scientifique.

L'autre souci que posent ces moteurs est la difficulté de trouver rapidement l'information dont on a besoin. On rappellera ici les quelques « trucs » essentiels pour trouver rapidement de l'information. D'abord, il faut se rappeler que les mots-clefs doivent être bien choisis : éviter les termes trop généraux et multiplier les mots-clefs pour affiner vos recherches. N'oubliez pas aussi d'utiliser les aides proposées par les sites pour améliorer vos recherches (<http://www.google.fr/intl/fr/help/basics.html>), tels les guillemets qui permettent de chercher une expression exacte.

Les moteurs ne sont pas les meilleurs outils pour débiter une recherche (les répertoires sont plus efficaces) mais ils peuvent devenir de vrai allié pour « dénicher » des objets rares ou invisibles sur des sites, à condition d'avoir déjà une idée précise de ce que l'on cherche.

b) Les répertoires

Pour débiter une recherche, mieux vaut passer par des répertoires de sites, qui font une sélection et ordonnent les sites par catégories. Il en existe de toutes sortes, depuis les portails les plus généralistes (<http://dmoz.fr/>) jusqu'au plus précis, le répertoire de lien de l'AJEI (<http://www.ajei.org/Netref/page/>) constitue un bon exemple, puisqu'il se concentre sur les sites ayant rapport aux sciences sociales et qui traitent du monde indien.

c) L'encyclopédie Wikipedia

Par ailleurs, il convient d'évoquer ici un autre outil, dont l'utilisation se répand, l'encyclopédie coopérative. Il s'agit d'un système qui permet à chaque internaute, sous le contrôle des autres, de constituer un savoir commun, articulé autour d'entrées, telle une encyclopédie classique. La plus connue, donc la plus fournie et la mieux construite, est le projet Wikipedia (<http://wikipedia.fr>, qui possède une version anglaise encore plus vaste : <http://en.wikipedia.org>). Elle constitue une ressource de qualité, mais qui n'a pas de comité de lecture scientifique et ne constitue donc pas un matériau que l'on peut directement utiliser et citer.

2. L'information para-scientifique

Internet a permis la mise en ligne et la diffusion de matériaux qui, s'ils n'ont pas le label de publication scientifique, n'en constituent pas moins des sources d'information intéressantes pour tout chercheur. Des fiches de lectures aux *working papers*, en passant par les rapports gouvernementaux¹⁷, un nombre croissant d'organisations et d'individus diffusent leur production via Internet. Ainsi, le *Center for International Development* de l'Université d'Harvard met en ligne les travaux de ses étudiants et une série de *working papers* (<http://www.cid.harvard.edu/cidwp>), de même que le Centre de Sciences Humaines de Delhi qui met librement à disposition ses *Occasional Papers* (<http://www.csh-delhi.com>).

Des projets de recherche ont désormais leur site dédié pour suivre au fur et à mesure le développement du programme de recherche ou rendre accessibles des ressources qui ne furent pas exploitées par les chercheurs pour une publication scientifique, tels les rapports, mémoires, voire bases de données. Ainsi, sur le site *demographie.net* (<http://www.demographie.net>) on trouve différentes rubriques, dont des cours donnés à l'EHESS, la production du *South Indian Fertility Project* ou encore des atlas sur l'Inde et l'annuaire des démographes de l'IRD. Créé et géré par un démographe de l'IRD, ce site n'est pas un site institutionnel, mais il fournit des éléments qui peuvent être cités dans d'autres travaux.

3. L'information scientifique

L'information scientifique, c'est-à-dire celle qui est approuvée par un comité de lecture constitué de scientifiques reconnus (*peer review* pour reprendre l'expression anglo-saxonne), est un « monde à part ». Elle est structurée par de grands éditeurs, des fondations, ou les institutions de tutelle (CNRS, universités, etc.).

Elle constitue le cœur du travail bibliographique scientifique.

La première source est l'ensemble des revues en ligne. On peut ainsi citer *Cybergeo*, revue européenne de géographie (<http://www.cybergeo.presse.fr>) ou la revue française *Mappemonde* (<http://mappemonde.mgm.fr/>). On trouve désormais d'autres revues en ligne ou dont les archives sont mises en ligne. On doit ici citer les plus importants projets que représentent JSTOR, ScienceDirect ou encore

Ingenta et le français revues.org. Derrière ces sites, se cachent des institutions scientifiques ou leurs membres (revues.org), des organisations à but non-lucratif (JSTOR) ou des maisons d'édition scientifique (Elsevier, Blackwell, Sage,...) qui valorisent leur catalogue de revue et numérisent leurs archives. De plus en plus de bibliothèques universitaires et de centres de documentation font appel à ce genre de service qui leur permet de résoudre le problème du stockage des revues. Si votre centre de recherche ou votre université souscrit à ces sites¹⁸, vous pouvez alors télécharger des articles de différentes revues, sur plusieurs années, souvent en format PDF. Dans le cas contraire vous devrez alors payer (et à plus de vingt euros l'article, cela fait réfléchir !).

Il existe une autre solution, dont le CNRS soutient activement le développement. Il s'agit du projet HAL (Hyper Article en Ligne) développé par le Centre pour la Communication Scientifique Directe (CCSD) et qui est un outil de communication scientifique directe entre chercheurs. L'idée est d'assurer gratuitement une meilleure diffusion des travaux de recherche. Les documents sont déposés par leurs auteurs, qui doivent veiller à leur bonne qualité scientifique. Des notices bibliographiques contenant les références d'une publication peuvent être mises en ligne, mais sans le texte intégral. On peut bien sûr mettre en ligne des textes complets, peu importe leur date de publication. Ce site est ensuite accessible à tous, via des entrées thématiques, disciplinaires ou un moteur de recherche dédié (<http://hal.ccsd.cnrs.fr/>).

Dans le même temps, le CCSD met aussi à disposition de la communauté scientifique un serveur de thèse (<http://tel.ccsd.cnrs.fr>), qui permet au doctorant de rendre son travail disponible, et aux autres chercheurs d'en disposer. Cette initiative s'inscrit dans la suite de projets lancés au Canada tels <http://www.collectionscanada.ca/thesescanada/> ou <http://www.cybertheses.org/> (coopération entre l'Université Lyon 2 et les Presses de l'Université de Montréal).

B. Restitution : La mise en ligne

Si l'exercice de recherche est jugé par les pairs, il ne paraît pas incongru, particulièrement aujourd'hui où l'on se pose la question de la place du chercheur dans la société, d'envisager de diffuser l'information au-delà de ce cercle des pairs. Internet offre cette possibilité, même si elle reste encore sous-exploitée.

1. Pourquoi mettre en ligne?

On peut trouver trois grands avantages à la mise en ligne : la visibilité, l'échange, le contrôle. Nous aborderons ces trois points, à travers l'exemple du site développé par Philippe Waniez, chercheur à l'IRD (<http://philgeo.club.fr>).

a) La visibilité

Il est parfois difficile de se faire connaître au-delà d'un certain cercle ou au contraire de suivre le parcours d'un chercheur et de ses projets. De la mise en ligne d'un curriculum vitae, à la présentation d'un projet de recherche et de ses résultats, Internet se pose comme une vitrine de plus en plus indispensable à l'activité de chercheur. Sur le site créé par Philippe Waniez on trouve non seulement son CV, mais aussi une revue de ses publications, de ses travaux de recherches et de ses projets passés et actuels (dont son Habilitation à Diriger des Recherches).

b) L'échange

C'est l'extension naturelle du point précédent. En étant plus visible, le chercheur peut ainsi encourager les échanges et les interactions avec d'autres personnes intéressées par ses projets. Il ne tient qu'au chercheur de se donner les moyens de nourrir la curiosité et l'intérêt de ces personnes, en mettant en ligne sa production et ses résultats.

Ainsi, P. Waniez a développé un logiciel de cartographie (Philcarto) qu'il a décidé de distribuer gratuitement à partir de son site. On trouve sur son site non seulement le logiciel, mais aussi l'aide nécessaire à son utilisation et plusieurs extensions (pour digitaliser des fonds de cartes ou convertir d'un format à un autre) ainsi que des données et des fonds de cartes utiles au fonctionnement du logiciel. Ce logiciel est aujourd'hui utilisé par les enseignants du secondaire et de certaines universités (Université Paris 1 Panthéon-Sorbonne, Université de Provence, par exemple).

A plusieurs reprises, P. Waniez demande aux visiteurs leur avis sur son logiciel ou ses travaux, avis que l'on peut rarement obtenir autrement dans le monde de la recherche (à l'exception de quelques revues d'ouvrages ou lors de colloques).

c) Le contrôle

Avec la mise en ligne de ses résultats de recherche ou de son projet, vient souvent la crainte de se faire copier ou de perdre le contrôle de ses résultats. Ces risques existaient déjà bien avant Internet, mais ce média parce qu'il donne l'impression d'être sans aucun contrôle ou sans aucunes limites dans son développement, attise les angoisses, telle celle d'être copié ou pillé.

Une autre manière de voir les choses serait au contraire de considérer la mise en ligne comme le meilleur moyen de garder le contrôle sur la diffusion de ses résultats. Qu'un lycéen un peu paresseux vous copie, ne portera pas grand préjudice à votre carrière de chercheur ou à vos projets. Qu'un spécialiste ou un chercheur mal intentionné reprenne votre prose sans vous citer et vous avez en ligne une preuve immédiatement communicable à tous de la faute déontologique dont vous venez d'être la victime. Au final, ce genre de désagrément apparaît mineur eu égard à la diffusion potentielle que peuvent atteindre vos travaux grâce à Internet.

Un mémoire de maîtrise qui dort quelque part sur une étagère de bibliothèque universitaire a effectivement moins de chance d'être copié qu'un mémoire en ligne. Il a aussi beaucoup moins de chances d'être lu... C'est pourquoi l'AJEI a mis en ligne depuis 1998 les communications de ses ateliers, séminaires et autres productions de ses membres (<http://www.ajei.org/fr/publications/publications.htm>). C'est pourquoi aussi les chercheurs ou les groupes multiplient aujourd'hui les sites personnels.

2. Comment mettre en ligne ?

Et si on franchissait le pas ? Utiliser les ressources de l'Internet est une chose, mettre à disposition son propre matériel en est une autre. Une fois la décision prise de mettre en ligne, encore faut-il trouver la solution la mieux adaptée à ses attentes.

Mettre en ligne nécessite de trouver un hébergeur qui va stocker et rediffuser votre site. Il existe des hébergeurs payants et d'autres gratuits. Les payants se classent en deux grandes catégories : les fournisseurs d'accès internet qui « offrent » de l'espace avec votre abonnement internet, et les « providers » dont le métier est de vendre de l'espace sur Internet. Au sein des hébergeurs gratuits on distinguera les institutions, gratuites mais qui ne laissent souvent que peu de place des hébergeurs, qui vivent de publicité qui surgissent sur votre site (pénibles pour les visiteurs).

a) Institutions

Aujourd'hui la majorité des centres de recherche actifs présentent des sites Internet offrant l'avantage d'être bien référencés par les moteurs de recherche généralistes et les répertoires, et donc d'avoir une bonne visibilité sur Internet. A défaut d'offrir de la souplesse dans la mise à jour de l'information, ils offrent une

bonne visibilité à leurs membres, qui peuvent lier leur propre site Internet depuis leur page de présentation (voir par exemple le site de l'UMR géographie-cités : <http://www.parigeo.cnrs.fr>).

Le Centre de Sciences Humaines de Delhi présente encore un bon exemple de mise à disposition de l'outil Internet pour ses membres. En effet, ce site met en ligne le CV, les publications, et les projets de recherche des membres du centre. Il constitue par ailleurs une mémoire du centre, conservant les informations sur le personnel qui n'y est plus présent¹⁹. Le chercheur a d'ailleurs tout intérêt à participer à ce genre de mise en ligne, puisque c'est de sa propre histoire dont il s'agit.

b) Initiative individuelle

(1) Compléter un site institutionnel

Face à la défaillance de leur institution de tutelle, ou plus simplement à cause du peu de possibilité offerte, certains chercheurs se tournent vers la mise en place de sites personnels. Sans revenir sur l'exemple de P. Waniez, on évoquera celui de l'historienne Inès Zupanov (membre du Centre d'Etudes de l'Inde et de l'Asie du Sud) qui lui permet en toute simplicité de mettre à disposition du public son CV détaillé (dont sa bibliographie complète), une partie de ses publications récentes ainsi que des images de ses terrains, notamment les églises goanaises (<http://www.inezsupanov.com/>). Le site web personnel permet de partager un peu plus que des documents strictement académiques. Le CEIAS, son centre de rattachement ne propose qu'une page assez courte sur ses chercheurs, mais offre un renvoi vers leur site personnel (<http://www.ehess.fr/centres/ceias/membres-disc.html>).

(2) Développer sa propre image

A une autre échelle, celle de la jeune recherche, l'intégration dans les laboratoires n'est pas toujours optimale, et développer son propre site permet de s'assurer une bonne visibilité, voire se donner un espace pour stocker son matériel de thèse en cours d'élaboration. Mieux qu'un disque dur, le web stocke vos données et surtout en rend une partie disponible, pour vos interlocuteurs. Face à la question récurrente : « mais vous faites quoi au juste ? », la réponse devient : « allez donc voir mon site. ». C'est ce qu'a fait par exemple Renaud Egreteau (<http://egreteau.com/>) doctorant en sciences politiques. Avant lui, P.A. Chouvy a créé un modèle du genre (<http://www.pa-chouvy.org>).

3. Comment mettre en ligne, concrètement ?

On ne parlera pas ici des modalités d'achat d'un nom de domaine (le nom du site) et d'espace de stockage en ligne, Internet regorge d'information, depuis la plus basique jusqu'à la plus technique. On rappellera par contre, qu'à partir d'une quinzaine d'euros on peut acheter son propre nom (www.abdeldupont.org par exemple), et qu'un site comme celui de l'AJEI coûte 45 euros par an. Un investissement qui reste mineur (6 roupies par jour autrement dit).

Pour mettre en ligne, plusieurs solutions s'offrent ensuite à vous : sites tout prêts (sites institutionnels notamment, mais aussi certains sites gratuits ou payants qui offrent des solutions toutes faites), construction de ses propres pages (avec un éditeur ou directement en code HTML - Hyper Text Markup Language), solutions plus avancées (php, spipe, etc.).

a) Les formats

Le langage de base pour créer des pages Web reste le HTML. Pour créer une page HTML nul besoin d'un logiciel de développement. Votre éditeur de texte²⁰ peut suffire. L'important étant de sauver votre page en format HTML et de mettre en ordre vos différentes pages par un système de liens.

En fait à peu près tous les formats de fichiers existants peuvent être mis en ligne. Les navigateurs Internet gèrent souvent très bien le lancement des logiciels nécessaires à la lecture d'un type de fichier particulier. Un

texte peut être lu ou téléchargé à partir d'un fichier Word (.doc) ou PDF. Les images (.gif, .jpeg, .bmp), le son (.wmv, .mp3), la vidéo (.mpeg, .avi, .wmv) sont aussi pris en charge par la plupart des navigateurs. On peut aussi développer des animations (type flash ou .svg) permettant de mélanger les différents formats et ainsi exploiter les potentialités multimédia offertes par le numérique (voir les présentations de Sébastien Oliveau et Bertrand Lefebvre dans ce même recueil).

Il s'agit ici de technologies plus avancées, mais qui restent accessibles à un jeune chercheur en sciences sociales qui ne souhaite pas devenir informaticien. Récemment, Pierre Chapelet et Bertrand Lefebvre ont ainsi proposé une version interactive de leur *Occasional Paper*²¹ contenant des cartes que l'on peut modifier (et s'appuyant sur la technologie Flash).

Dans le même ordre d'idée, l'atlas interactif d'Anne-Claire Couïc, disponible sur le site www.geodemo.net, offre des possibilités variées d'exploration cartographique du territoire indien, en s'appuyant sur des technologies accessibles sans programmation.

b) Les solutions pour créer ses pages

Une fois que l'on a trouvé son hébergeur il faut créer son site. Certains vous diront que rien ne vaut la programmation HTML, mais tout le monde n'y a pas forcément intérêt (faut-il un BEP mécanique-auto pour conduire ?). Une autre solution est l'utilisation de logiciels WIZWIG (*what you see is what you get*) comme Dreamweaver (de la société Macromédia ou Golive de la société Adobe), qui s'utilisent comme des logiciels de traitement de texte. Mais cette solution a un coût important, et n'est pas toujours non plus la plus simple (les logiciels ne se ressemblent pas, etc.)

Il existe des solutions gratuites de mise en forme automatique (le standard actuel étant SPIP), qui permettent une gestion aisée de son site, à condition toutefois de passer le cap de sa mise en marche initiale. SPIP inclut plusieurs automatismes qui vont vous permettre de partager la gestion d'un site, de créer vos pages Web sans rien connaître du langage HTML ou encore de modifier facilement l'architecture de votre site. Vous trouverez beaucoup plus d'informations sur le site du projet (<http://www.spip.net/fr>). Un exemple d'interface simple et efficace, sous SPIP, est offert par le site de Gilles Verniers, doctorant à l'IEP de Paris qui donne accès à l'actualité politique indienne (<http://india.eu.org/>).

C. Pour conclure

Ce très bref tour d'horizon avait pour but d'éclairer rapidement les possibilités et les enjeux basiques de l'Internet dans la recherche contemporaine, en insistant sur des réalisations centrées sur la recherche en sciences sociales sur le monde indien. Le lecteur curieux aura tout intérêt à visiter les différents sites qui lui permettront de mieux comprendre notre propos.

Enfin, on se doit de rappeler que le site de l'AJEI (<http://www.ajei.org>) offre une première marche pour la mise en ligne de mémoires et d'articles (mais aussi de photos), et offre via ses forums une information large et un lieu pour poser ses questions.

Le lecteur qui souhaiterait aller plus loin, peut contacter les auteurs ou plus simplement chercher via les moteurs de recherche et autres répertoires les sites et forums qui les aideront à démarrer. L'Internet regorge de bonnes âmes dévouées à la mise en ligne qui sauront les guider dans leurs premiers pas.

Notes

¹⁵ Actuel webmaster de l'AJEI, secrétaire scientifique du CSH : <http://csh-delhi.com/team/team.php?idstaff=101>.

¹⁶ Ancien webmaster de l'AJEI, webmaster du nouveau site de l'UFR de sciences géographiques et de l'aménagement de l'université de Provence où il est maître de conférences. Webmaster du site <http://www.geodemo.net>.

¹⁷ Ceux des agences de l'ONU sont systématiquement mis en ligne, par exemple les rapports sur le développement humain : <http://hdr.undp.org/>.

¹⁸ D'où la nécessité de se renseigner – les laboratoires CNRS y ont accès – et de faire pression sur ses laboratoires et universités pour y être abonné.

¹⁹ Les postes de chercheurs dans les centres français de recherche à l'étranger sont des contrats de quelques années, ce qui implique un *turnover* important du personnel scientifique, et souvent un problème de visibilité et de continuité, qu'un site Internet avec des archives peut partiellement combler.

²⁰ Si la plupart utilise le logiciel Word de Microsoft, on doit rappeler l'existence d'un concurrent très sérieux et gratuit, Openoffice, dont les fonctionnalités Web sont bien meilleures (<http://fr.openoffice.org/>).

²¹ Chapelet Pierre, Lefebvre Bertrand, (2005), *Contextualizing the Urban Healthcare System. Methodology for developing a geodatabase of Delhi's healthcare system*, CSH Occasional Paper N°11, Publication of the French Research Institutes in India, Rajdhani Art Press, New Delhi. Disponible en ligne: <http://csh-delhi.com/electronicop11/start.htm>